

Analisis Sentimen Ulasan *Video Game* Steam Disco Elysium Menggunakan Metode *Random Forest* dengan Fitur *Chi-Square*

Dendy Hadinata¹, Mahendra Dwifabri Purbolaksono², Utami Kusuma Dewi³

^{1,2}Fakultas Informatika, Universitas Telkom, Bandung

¹dendyhadinata@student.telkomuniversity.ac.id, ²mahendradp@telkomuniversity.ac.id,

³utamikusdew@telkomuniversity.ac.id

Abstrak

Video game telah muncul sebagai salah satu bentuk hiburan paling populer di seluruh dunia, terutama melalui platform distribusi digital seperti Steam yang menyediakan akses ke lebih dari 30.000 game. Ulasan pengguna di Steam sangat penting bagi pengembang game untuk memahami kebutuhan konsumen dan meningkatkan kualitas produk. Penelitian ini bertujuan untuk melakukan analisis sentimen terhadap ulasan pengguna pada game Disco Elysium di Steam dengan menggunakan metode klasifikasi Random Forest yang dioptimalkan melalui seleksi fitur Chi-Square. Dataset yang digunakan mencakup 5.000 ulasan yang telah diberi label secara manual, dikategorikan sebagai sentimen positif atau negatif. Langkah preprocessing akan dilakukan untuk mempersiapkan data mentah menjadi data yang siap diolah pada klasifikasi model. Metode TF-IDF kemudian diterapkan untuk pembobotan kata, diikuti oleh Chi-Square untuk seleksi fitur guna mengurangi fitur yang tidak relevan. Model Random Forest diterapkan dengan penyetelan hyperparameter menggunakan GridSearchCV dan RandomizedSearchCV. Matrik evaluasi, termasuk akurasi, presisi, recall, dan skor F1, dihitung menggunakan confusion matrix. Hasil penelitian menunjukkan bahwa stemming dan penyetelan hyperparameter meningkatkan akurasi model, sementara seleksi fitur Chi-Square tidak selalu memberikan peningkatan kinerja yang signifikan. RandomizedSearchCV mengungguli GridSearchCV dalam hal efisiensi komputasi dan eksplorasi parameter optimal, sehingga lebih efektif untuk kebutuhan penelitian ini. Penelitian ini menyoroti potensi optimalisasi lebih lanjut dalam teknik preprocessing dan seleksi fitur untuk meningkatkan analisis sentimen berbasis teks. Hasil dari penelitian ini mendapatkan nilai tertinggi pada model RandomizedSearchCV dengan hasil 88,7% sedangkan GridSearchCV 88,6%, menunjukkan bahwa hasil modelling dari kedua metode tersebut memiliki hasil skor yang optimal. Temuan dari penelitian ini diharapkan dapat membantu pengembang game dalam memahami sentimen pengguna dari ulasan di platform digital secara lebih efisien, sehingga mendukung pengambilan keputusan dan peningkatan produk yang lebih terinformasi.

Kata kunci : *Analisis Sentimen, Steam, Disco Elysium, Random Forest, Chi-Square*

Abstract

Video games have emerged as one of the most popular forms of entertainment worldwide, particularly through digital distribution platforms like Steam, which provides access to over 30,000 games. User reviews on Steam are crucial for game developers to understand consumer needs and improve product quality. This study aims to conduct sentiment analysis on user reviews for the game Disco Elysium on Steam using the Random Forest classification method, optimized through Chi-Square feature selection. The dataset comprises 5,000 manually labeled reviews categorized as positive or negative. Preprocessing steps include punctuation and number removal, lowercasing, stopword elimination, stemming, and tokenization. The TF-IDF method is then applied for word weighting, followed by Chi-Square for feature selection to reduce irrelevant features. The Random Forest model is subsequently applied with hyperparameter tuning via GridSearchCV and RandomizedSearchCV. Evaluation metrics, including accuracy, precision, recall, and F1-score, are calculated using a confusion matrix. Results indicate that stemming and hyperparameter tuning enhance model accuracy, while Chi-Square feature selection does not always significantly improve performance. RandomizedSearchCV outperforms GridSearchCV in terms of computational efficiency and optimal parameter exploration, proving more effective for this study's requirements. This research highlights the potential for further optimization in preprocessing techniques and feature selection to enhance text-based sentiment analysis. The results of this study indicate that the *RandomizedSearchCV* model achieved the highest score of 88.7%, while *GridSearchCV* obtained 88.6% this suggests that modeling using both methods produces optimal scores. Findings from this study can guide game developers in efficiently understanding user sentiment from digital platform reviews, thus aiding in more informed product improvement and decision-making.

Keywords: Sentiment Analysis, Steam, Disco Elysium, Random Forest, Chi-Square
